Summary of Revision

To the entire review team:

Thank you for a positive and thoughtful evaluation of our paper, we were grateful to receive constructive comments from the reviewers. In response, we have made a number of substantial improvements to the paper. We first summarized the major changes below and then address each reviewer's comments individually.

1. **Explanation of Markdown Dimension.** Some reviewers believe that the explanation of our key concept, markdown dimension, can be improved. We added one subsection (Subsection 2.2 in the revised version) that analyzes the nice properties of the linear demand family ("robustness under noisy demands") and used it to motivate more general markdown dimensions.

2. **Monotonicity of Prices in Algorithm 1.** We thank reviewers ZpZo and CjUD for pointing out that in Algorithm 1 the prices are not guaranteed to be decreasing. This can be fixed by a slight modification. We added Line 12 in Algorithm 1, which ensures monotonicity: it compares the estimated optimal price $\tilde{p}_{j+1} = \arg\max\{p^*(\theta) : |\theta - \theta^*| \leq w_{j+1}\}$ (called the "raw" price), with the previous price p_j.

   We show, however, that such modification does not lead to significant change in the analysis. In fact, we show that with $1 - O(T^{-2})$ probability, we have $\tilde{p}_{j+1} \leq p_j$. Thus, by losing an $O(T) \cdot O(T^{-2}) = o(1)$ factor in the regret (due to the low probability event), our **previous** analysis goes through, that is, it suffices to just analyze the regret for the price sequence $\{\tilde{p}_{j+1}\}$. We have revised the analysis for $d = 0$ accordingly in the supplementary materials.

3. **Missing Proofs.** Reviwer CjUD pointed out that the proofs of Propositions 1,2,3,4 are missing. While the proofs are actually fairly straightforward, we agree with the reviewers that adding them would make it easier for readers to digest the definition of markdown dimension, as well as improving the completeness of the paper. We have therefore added their proofs in the revised supplementary materials.

4. Experiments. Some reviewers suggested that numerical simulations would significantly improve this work. This is a good suggestion, but unfortunately, due to the large volume of other issues to address, we were not able to prioritize the experiments in the rebuttal period, but we will add a comprehensive experiment in the camera-ready version if this work is accepted.

Reviewer ZoQT

Thank you for giving us the chance to revise the paper. We deeply appreciate the time you spent on the paper and the detailed feedback that you have given us.

1. *"the main weakness of the paper is that individual algorithms are presented for different demand markdown-dimension regimes, meaning that the decision maker would need to know whether the supposedly unknown demand function's markdown dimension is 0, finite or infinite, in order for a decision maker to deploy the corresponding algorithm. Also, in the finite, non-zero markdown dimension regime, the proposed Algorithm 2 requires the decision maker to know the underlying markdown dimension of demand. The paper would be much stronger if it can either propose a single "best of all worlds" algorithm that can achieve optimal regret under demands with any markdown dimensions while being agnostic to the regime, or analyze how the algorithms proposed in the papers would perform if the regime is misspecified."*

   We add a remark in the end that says if the learner only knows an upper bound $d_{\max}$ on the true dimension d, then she can achieve the guarantee for $d_{\max}$ by simply running the optimal policy for $d_{\max}$-dimension family. In particular, if the learner knows nothing about the dimension, i.e. $d_{\max} = \infty$, then she may simply choose the policy in Chen NeurIPS '21 or Jia et al 2021 to ensure $T^{3/4}$ regret.

2. *"The two key characteristics for demand functions in this paper are the markdown dimension and sensitivity. However, I noticed that there is no discussion on how these parameters collectively affect regret. For example, in Theorem 1 and 2, is there any intuition behind why there is a discontinuous "jump" in regret for $s = 2$ for 0 markdown dimension, but for finite markdown dimension such a "jump" disappears?"*

   It is a good suggestion. We have added several remarks below the results involving both d and s. As we recall, the main factor in the regret are caused by "overshooting", i.e., stopping at a price lower than the optimal price. If s is large, the function changes at a slower rate around the optimum, and hence the policy suffers less, therefore one should expect the regret to improve. This intuition matches our upper bounds for all three regimes.

Regarding the "jump" from $\log T$ to $\log^2 T$ in the $d = 0$ case: this is essentially caused by jump in s of the integral $\int_0^{\log T} c^{(1-s/2)x} dx$. In fact, when $s = 2$, the integrand is constant and hence the integral is $\sim \log T$. When $s > 2$, the integrand becomes an exponentially decreasing function, and so the intergal is $O(1)$. Such jump in the exponent of the $\log T$ term do not appear in the bounds for $d \geq 1$ and $d = \infty$ since they are using different policies. To sum up, our results suggest that s affects the exponent of T continuously, but may affect the exponent of $\log T$ in a discontinous way.

3. *"In Algorithm 2 and Theorem 3, how should the decision maker choose h and the phase lengths $T_1, ..., T_m$?*

   In the revised version we specified the optimal choice of T_j's and h in both the main body and the supplementary. For $s = 2$, the optimal T_j is $T^{\frac{md+j}{(m(d+1)+1}}$ for each $j = 1, ..., m$, the optimal h is $T^{\frac{m}{(m(d+1)+1}}$, and the order-optimal choice of m, the number of phases, is $\log T$.

4. *"This suggestion may be beyond the scope of the paper, but perhaps it would be helpful to run some simulations to shed light on relevant aspects."*

   This is a good suggestion. Unfortunately, due to the large volume of other issues to address, we were not able to prioritize the experiments in the rebuttal period, but we will add a comprehensive experiment in the camera-ready version if this work is accepted.

Reviewer Vvni

Thank you for giving us the chance to revise the paper. We deeply appreciate the time you spent on the paper and the detailed feedback that you have given us.

1. *"The main weakness is that the setting is not convincingly motivated: Since the markdown constraint is motivated by pricing problems in markets of some sort, I am not sure it is reasonable to assume that the form of demand (essentially how the market will react to the price) is known. "*

   It is in fact commonly assumed in the revenue management community that the demand function may take certain functional form. The following work explained why it is reasonable to assume parametric forms, **specifically for markdown pricing**:

   - Smith, Achabal, Clearance pricing and inventory policies for retail chains, Management Science '98
   - "Markdown Managment", Rama Ramakrishnan, The Oxford Handbook of Pricing and Revenue management

2. There is no concrete example where a parametric form of the demand model is reasonable/accurate and no numerical results to illustrate the performance of the proposed algorithms vs standard algorithms in practical problem settings. It would be nice to provide numerical results with real datasets illustrating the performance of the proposed algorithms by fitting/assuming different kinds of models for the data.

   This is a good suggestion. Unfortunately, due to the large volume of other issues to address, we were not able to prioritize the experiments in the rebuttal period, but we will add a comprehensive experiment in the camera-ready version if this work is accepted.

Reviewer CjUD

Thank you for giving us the chance to revise the paper. We deeply appreciate the time you spent on the paper and the detailed feedback that you have given us.

1. The major question is about Algorithm 1 (on page 7). Is there any step in Algorithm 1 to make sure the prices pj are decreasing? More precisely, in phase j the parameter j is estimated, and the estimation may be inaccurate with a probability. The conservative price p_{j+1} (line 11) obtained by the inaccurate estimation may be larger than p_j. After checking the algorithm and proof, I have not found any steps to prevent this. good catch.

   This is a good point. This can be fixed by a slight modification to Algorithm 1 as well as the analysis. In the revision, we added in Line 12 in Algorithm 1 which ensures monotonicity: it compares the estimated optimal price $\tilde{p}_{j+1} = \arg\max\{p^*(\theta) : |\theta - \theta^*| \leq w_{j+1}\}$ (which we called "raw" price), with the previous price p_j.

   We show, however, that such modification does not lead to significant change in the analysis. In fact, we show that with $1 - O(T^{-2})$ probability, we have $\tilde{p}_{j+1} \leq p_j$. Thus, by losing an $O(T) \cdot O(T^{-2}) = o(1)$ factor in the regret (due to the low probability event), our **previous** analysis goes through, that is, it suffices to just analyze the regret for the price sequence $\{\tilde{p}_{j+1}\}$. We have revised the analysis for $d = 0$ accordingly in the supplementary materials.

2. *"The regret for the parametric case when $d = \infty$, does not match the regret $O(T^{3/4})$ for the nonparametric unimodal case. So after imposing the parametric assumption (when $d > 3$), the regret may be worse. Yet it is shown that it matches the lower bound. Does it make the definition of unnatural?"*

   Good question, but this is not a contradiction. In fact, the $T^{3/4}$ bound assumes unimodality, which may not hold for d-dimensional family. For example, when $d \geq 4$, degree-d polynomial is d-dimensional but in general not unimodal.

   As a side note, observe that when $d = 1$ (i.e. linear demand), the reward function is quadratic and hence unimodal, so the $T^{3/4}$ bound applies. In this case, our bound

becomes $T^{1/2}$, strictly better than the $T^{3/4}$ bound for **general** unimodal reward function. We have added this discussion after Theorem 3.

3. *"Also, I have not found the proof for Proposition 1,2,3,4 in the paper or appendix."*

   We have added their proofs in Appendix A in the revised supplementary.

Reviewer ZpZo

Thank you for giving us the chance to revise the paper. We deeply appreciate the time you spent on the paper and the detailed feedback that you have given us.

1. *"By the end of the first paragraph of Section 2: What do you mean by "almost surely"? As T is finite, an "almost surely" constraint means no violations at all, right?"*

   This is a just a rigorous way of saying some event holds with probability 1, in other words, the event of violation has probability measure 0 (though it may be non-empty). We have replaced the phrase "almost surely" with "with probability 1".

2. *"In the second paragraph of Section 2: It is better to define a policy π formally, as it is time-and-history dependent."*

   We added a formal definition of "policy" as a footnote in page 3.

3. *"Assumption 1: Why the optimal price can be assumed to be an interior point instead of on the boundary, without loss of generality?"*

   We did not claim this is "without loss of generality". In fact, it does lose generality. While this interior point assumption is quite common in previous literature, after revisiting our proofs, we realized that it is actually possible to remove Assumption 1 and and extend our results even further, without significantly modifying the analysis.

   More precisely, recall that with Assumption 1, we have that $s \geq 2$ for twice-differentiabile family (due to Taylor expansion and vanishing first derivative at the optimum). Now, without Assumption 1, the sensitivity s can possibly be 1. By plugging in s=1 in our analyses, we obtain the following results:

   - Alg 1 achieves $\sqrt{T}$ regret for 0-dimensional family when $s = 1$. This bound is higher than the upper bound for $s = 2$, which highlights the impact of s.
   - Alg 2 achieves $T^{(2d+1)/(2d+2)}$ regret for finite $d \geq 1$ dimensional family with $s = 1$. This bound is again higher than the corresponding upper bound for $s = 2$.

- Moreover, infinite dimensional family with $s = 1$ is already studied in Chen '21 and Jia et al '21, with a tight $T^3/4$ upper bound. This exactly matches our $T^{(2s+1)/(3s+1)}$ bound with $s = 1$.

We will double check the analysis carefully, and will add the above results to the camera ready version if accepted.

4. *"The first equation on the top of Page 4: With only differentiable assumption, there does not necessarily exist such a second-order derivatives. Notice that a twice-differentiable assumption is in Assumption 4 Therefore, I suggest the authors to state Assumption 1 more clearly."*

   This is a good point, we haven't assumed twice-differentiability until Assumption 4. In the revised version we added at the top of page 4 that "in addition, if R is twice differentiable".

5. *"One line before Assumption 5: I suggest the authors to formally define the p^* function in formulas."*

   We added a formal definition right before Assumption 5.

6. *"Assumption 5: p^* is a function mapping the parameter of D to the optimal price of D. On the one hand, the property of p^* is dependent on that of D, especially on how a θ determines a $D(p;\theta)$. Also, p^* is not an explicit function under most circumstance, but D is obviously more tractible. So I think it is not a proper way to make assumptions on p^* directly."*

   This is actually a quite common assumption in the literature of parametric demand learning, e.g. see Assumption 1(c) of "Dynamic Pricing Under a General Parametric Choice Model" (OR'12), by Broder and Rusmevichientong. Moreover, it is satisfied by many commonly used demand functions such as linear, exponential and logit demand. In fact, in these cases, p^* admits a **closed form** expression (see our proof for Proposition 3 in the Appendix), which can be easily verified to be Lipschitz on suitable domain.

   That said, we agree that such an assumption is somewhat indirect. However, we do not think it would be easy to circumvent this issue. In fact, take the family polynomial

demand functions as an example. While one can show that p^* is continuous in the coefficients, it is more challenging to show Lipschitzness, even though the domain is compact. We have added the discussion on this limitation in the revised version right after we Assumption 5.

7. *"Proposition 4: Is this easy to be proved?"*

   Yes, we have added some discussion right after Proposition 4. To show a family has infinite dimension, it suffices to show that for any finite integer d, this family is **not** d-identifiable. In other words, there exists d distinct prices, and two distinct functions having the same values on those d prices. To see this, for the family $\mathcal{F}$ of 1-Lipschitz functions on $[0, 1]$, consider $f(x) = \sin \frac{d\pi x}{d\pi}$ and $g(x) = \sin \frac{2d\pi x}{2d\pi}$ for any given d. Then, f and g are both 1-Lipschitz, and $f(\frac{i}{d}) = g(\frac{i}{d}) = 0$ for any $i = 0, 1, ..., d$, so $\mathcal{F}$ is not d-identifiable.

8. *"The last paragraph on Page 6: What is C_s? (I guess it is the s-th coefficient of the p^s term of R's polynomial approximation, but not sure if I am right.) Besides, this inequality is not always true as you neglect the Lagrangian remainders."*

   This can be fixed by adding an assumption (item (d) in Definition 10 of the revised version) that the j-th derivative of any reward function is bounded by C_j for $j = 0, 1, 2, ..., s$. We defined this only for $j = 0, 1, 2$ in Assumption 4 but forget to do so for $j > 2$.

   With this definition, the inequality then follows. In fact, by Taylor's theorem with Lagrangian remainder, for any $\varepsilon > 0$, there exists some value z between x^* and $x^* + \varepsilon$ such that

   $$R(x^*) = R(x^* + \varepsilon) + 0 + ... + 0 + \frac{1}{s!} \cdot R^{(s)}(z)\varepsilon^s.$$

   Since $|R^{(s)}(z)| \leq C_s$, we have $|R(x^*) - R(x^* + \varepsilon)| \leq \frac{1}{s!} C_s \varepsilon^s$.

9. *"The first line on Page 7: It is NOT what [10] have stated. Their definition on α (albeit the s here) is a Lipschitz exponent instead of a sensitive parameter. In other words, they aim at upper bound the difference while sensitivity is a lower bound of difference. This might be caused by a mistake in stating Definition 10 (i.e., C_6 serves as a lower*

bound instead of an upper bound), but I got really confused about its name "sensitivity". Intuitively, a larger sensitivity would cause a larger fluctuation on reward (or demand) while changing the price, but a larger s would definitely "smoothen" the demand curve and reduce the fluctuation caused by price changing. I am very curious why the authors did not call it a "smoothness"? Maybe there exist some causes in depth, and I am also not sure if this would affect the soundness."

We realized that the local Lipschitzness in Kleinberg '04 is actually **not** comparable to our notion of sensitivity. Recall that in Kleinberg '04, a function $f(x)$ is "α-locally Lipschitz" for some $\alpha \in [0, 1]$, if there exists constants $L, \delta > 0$ and for any x, y with $|x - y| < \delta$, it holds that $|f(x) - f(y)| \leq L|x - y|^\alpha$. Further, for $\alpha > 1$, this class of functions is actually trivial. In fact, one can verify that only constant functions satisfy the above inequality.

In contrast, our sensitivity s only involves the "neighborhood" of p^*, but not other regions. Also, by our definition, the discussion on sensitivity focuses on $s \geq 1$, and in fact it does not make sense to consider $s < 1$. Therefore, we should not compare the results for our sensitivity with that of Kleinberg's local Lipschitz exponent. We have removed the paragraph involving such comparison, and revised other relevant places.

10. *"Algorithm 1 Line 11: How to ensure $p_{j+1} \leq p_j$ as θ^j is a random variable?*

This is a good point. This can be fixed by a slight modification to Algorithm 1 as well as the analysis. In the revision, we added in Line 12 in Algorithm 1 which ensures monotonicity: it compares the estimated optimal price $\tilde{p}_{j+1} = \arg\max\{p^*(\theta) : |\theta - \theta^*| \leq w_{j+1}\}$ (which we called "raw" price), with the previous price p_j.

We show, however, that such modification does not lead to significant change in the analysis. In fact, we show that with $1 - O(T^{-2})$ probability, we have $\tilde{p}_{j+1} \leq p_j$. Thus, by losing an $O(T) \cdot O(T^{-2}) = o(1)$ factor in the regret (due to the low probability event), our **previous** analysis goes through, that is, it suffices to just analyze the regret for the price sequence $\{\tilde{p}_{j+1}\}$. We have revised the analysis for $d = 0$ accordingly in the supplementary materials.

11. *"In Theorem 3: It might be better to rewrite the expression for $s = 2$ and $s > 2$, which would make both of them concise and informative."*

    We have rephrased Theorem 3 accordingly.

Reviewer WcPP

Thank you for giving us the chance to revise the paper. We deeply appreciate the time you spent on the paper and the detailed feedback that you have given us.

1. *"It might need more intuitions or explanations when introducing the new notion markdown dimension. In fact, it would be much better to first consider some special cases and then introduce the general notion based on the intuitions from the special cases. This will help readers appreciate the contributions."*

   We added one subsection (subsection 2.2 in the revised version) that analyzes the nice properties of the linear demand family ("robustness under noisy demands") and used it to motivate more general markdown dimensions.

2. *"The current version does not contain any experiments. The reviewer thinks that this is even more necessary for this paper since the goal is to show some improvement under each particular function class. Thus, it would be good to see improvements from experiments, at least some simple numerical experiments."*

   This is a good suggestion. Unfortunately, due to the large volume of other issues to address, we were not able to prioritize the experiments in the rebuttal period, but we will add a comprehensive experiment in the camera-ready version if this work is accepted.

3. *"Some possible typos: (1) the first line on Page 2, extra 'this bound'? (2) Proposition 3, b should be a? "*

   Fixed.

References

[1] R. Agrawal. The continuum-armed bandit problem. *SIAM journal on control and optimization*, 33(6):1926–1951, 1995.

[2] P. Auer, R. Ortner, and C. Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. In *International Conference on Computational Learning Theory*, pages 454–468. Springer, 2007.

[3] J. Broder and P. Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012.

[4] N. Chen. Multi-armed bandit requiring monotone arm sequences. *Advances in Neural Information Processing Systems*, 34, 2021.

[5] W. C. Cheung, D. Simchi-Levi, and H. Wang. Dynamic pricing and demand learning with limited price experimentation. *Operations Research*, 65(6):1722–1731, 2017.

[6] A. Garivier, P. Ménard, L. Rossi, and P. Menard. Thresholding bandit for dose-ranging: The impact of monotonicity. *arXiv preprint arXiv:1711.04454*, 2017.

[7] W. Gautschi. On the inverses of vandermonde and confluent vandermonde matrices. i, ii. *Numer. Math*, 4:117–123, 1962.

[8] S. Gupta and V. Kamble. Individual fairness in hindsight. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 805–806, 2019.

[9] S. Jia, A. Li, and R. Ravi. Markdown pricing under unknown demand. *Available at SSRN 3861379*, 2021.

[10] R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 594–605. IEEE, 2003.

[11] R. D. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pages 697–704, 2005.

[12] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[13] M. Luca and O. Reshef. The effect of price on firm reputation. *Management Science*, 2021.

[14] G. Perakis and D. Singhvi. Dynamic pricing with unknown non-parametric demand and limited price changes. *Available at SSRN 3336949*, 2019.

[15] J. Salem, S. Gupta, and V. Kamble. Taming wild price fluctuations: Monotone stochastic convex optimization with bandit feedback. *arXiv preprint arXiv:2103.09287*, 2021.

[16] D. Simchi-Levi and Y. Xu. Phase transitions and cyclic phenomena in bandits with switching constraints. *Advances in Neural Information Processing Systems*, 32, 2019.

[17] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[18] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339, 1948.